

Targeted Resequencing, the *Genotypic* Experience:

from custom Probe/ Bait design, DNA target capture with Agilent "SureSelect™" (Array-based and Liquid-phase Capture including Multiplexing), Illumina's NGS platform & Bioinformatics capabilities for Sequence Data Analysis.

Raja Mugasimangalam, Deepak S A, Shruti Sinha, Rohit Shukla, Veer Marwah, Vasanthan J, Anand C Damodaran.

www.genotypic.co.in
genomics@genotypic.co.in

Genotypic Technology Pvt. Ltd , #259, "Aparva", 4th Cross, RMV 2nd Stage, 80 feet Road, Bangalore-560094, India

Next/ Now Generation (NGS) technologies allow ultra-high deep sequencing of a given genomic region and to identify and validate genomic variations with a higher degree of precision in a shorter period of time. Robust methods have been developed in parallel for DNA target capture/ enrichment/ selection of specific genes, genomic regions, exons, whole exome etc. The complementary technologies provide scientists a powerful and flexible platform for Targeted Resequencing to focus on the regions of their interest. These methods are quicker, less expensive, simpler and offer higher resolution at a given location than traditional multiple-PCR based enrichment followed by Sanger Sequencing.

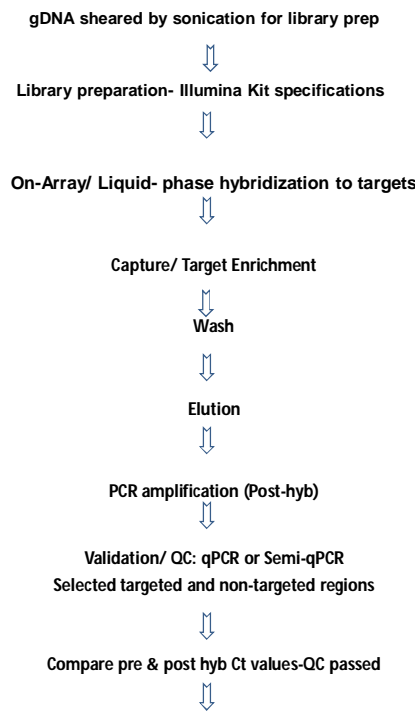
We have completed targeted resequencing of human, plant and animal genomes with improvements and optimizations in all steps. Our pipeline includes **Illumina NGS platform (GAIIx) and Agilent "SureSelect DNA capture arrays"(On-array/ solid-phase) and "SureSelect Target enrichment"(liquid phase) methods**. We have optimized multiplexing in the array based Capture method using PCR-based barcodes and also proprietary ligation-based barcoding methods. Custom Capture Probe/ Bait design for targeted regions is done using Agilent e-Array, Sequence Data Analysis for Variation/ Mutation/ SNP discovery through in-house bioinformatics capabilities.

CUSTOM PROBE/ BAIT LIBRARY DESIGN
Target regions- specific chromosomal intervals, coordinates, list of genes, RefSeqs, Exons, whole exome
Agilent e-Array tool used; Repeat Regions Masked
50 -100 bp flanking regions of introns in exon boundareis also covered for efficient capture
Deliverable -Capture microarray design file (.bed file loadable to genome browsers), probe coverage and quality control report
Capture design also possible from cDNA sequence and partial genomic sequence to capture from genomic DNA

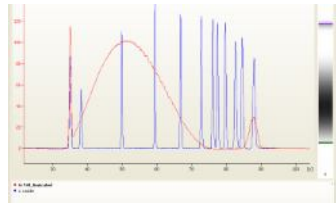
Genotypic Catalog Designs for On-Array DNA Capture
✓ Human Whole Exome (Exons of whole genome)
✓ Human Hepatic drug metabolism genes
✓ Human and Mouse miRNA promoter
✓ Designs also available for Rat, Guinea pig, other mammalian and plant species.

DNA CAPTURE using AGILENT "SURESELECT" On-Array; In-Solution
gDNA or prepared libraries is starting material. Library preparation for Illumina NGS using Illumina kit (singleton, paired end)
On-Array Capture - 6 Mb target can be captured using Agilent 244k wit th DNA probes- hybridization on-array
In-solution/ Liquid-phase- 3. 3 Mb captured at 2x tiling; 6.6 Mb captured at 1 X tiling
Bait libraries (RNA) hybridized to target in-solution/ liquid phase capture
Multiplexing up to 6 samples per Capture validated for On-Array capture and brings down costs
BARCODES/ TAGS/ INDEX: Illumina's standard barcodes (PCR primer based) and ligation based barcodes. Use of PCR based barcodes have shown dimers requiring clean up
Genotypic was a beta testing site for On-Array Capture (2008) and methodology is optimized and robust

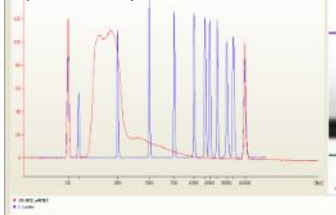
On-ARRAY Capture / In-Solution Capture Workflow



Sheared genomic DNA- Sonication (pre-hyb)



Captured Library Post Enrichment

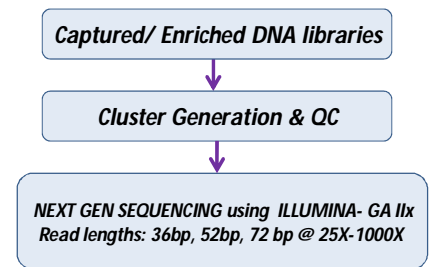


Performance – Agilent Capture method tested at Genotypic

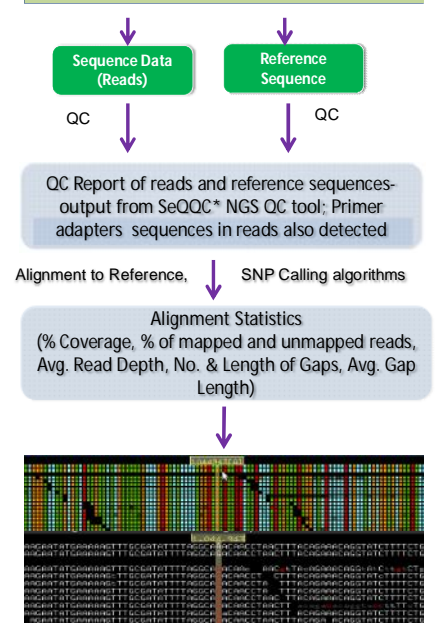
Enrichment of targeted regions (qPCR)	~1000 Fold
Depletion of untargeted regions (qPCR)	~400 Fold
Sequence Reads mapping to targeted regions	66. 85 %
Targeted regions enriched	99 %
Reads mapping back to Targeted regions	50- 60 %

NEXT GEN SEQUENCING- ILLUMINA GAIIX PLATFORM SEQUENCE DATA ANALYSIS- VARIATION DISCOVERY

Sequencing done primarily using 54 bp singleton runs; also 36 bp singleton and 72 bp singletons, 72 bp paired end runs



Sequence Data Analysis @ Genotypic Variation Discovery



SNP/ Mutation Discovery Reports with Annotation

Reference Sequence	Genomic Position	Reference Base	Observed Base	Strand	Chromosome	Gene_ID	CCDS_ID	GeneStart	GeneEnd	ExonCount	ExonStarts	ExonEnds
chrX:129229683-139883213	130243499	T	Y	-	chrX	3547	CCDS14629.1	130235450	130248329	19	130235450,1	130235586,1
chrX:129229683-139883213	130244639	C	T	+	chrX	3547	CCDS14629.1	130235450	130248329	19	130235450,1	130235586,1
chrX:129229683-139883213	134775700	A	G	-	chrX	441521	CCDS35406.1	134773822	134777912	4	134773822,	134773880,1

SNPs matched to dbSNP and Hapmap database

Reference Sequence	Genomic Position	Reference Base	Observed Base	ReadDepth	Zygoty	RS_ID	Strand	SNP_in_DBSNP	SNP_in_CEU	SNP_in_CHB	SNP_in_JPT	SNP_in_YRI
chrX:129229683-139883213	130243499	T	Y	16	T(9)/C(7)	rs1128617	-	A/G	A/G	A/G	A/G	A/G
chrX:129229683-139883213	130244639	C	T	26	Homozygous	rs5932877	+	C/T	C/T	C/T	C/T	C/T
chrX:129229683-139883213	134775700	A	G	32	Homozygous	rs2034920	-	C/T	C/T	C/T	C/T	C/T